DOCUMENT RESUME

ED 290 767 TM 011 007

AUTHOR Lancaster, Diana M.; And Others
TITLE A Comparison of Item Type and Source on Difficulty and Discrimination Ability.
PUB DATE Nov 87
NOTE 12p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (16th, Mobile, AL, November 11-13, 1987).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Dental Schools; Dentistry; Difficulty Level; *Discriminant Analysis; Higher Education; *Item Analysis; *Multiple Choice Tests; Statistical Analysis; *Test Construction; Test Format; *Test Items; Test Theory
IDENTIFIERS Item Discrimination (Tests); Louisiana State University; Short Answer Tests

ABSTRACT

Difficulty and discrimination ability were compared between multiple choice and short answer items in midterm and final examinations for the internal medicine course at Louisiana State University School of Dentistry. The examinations were administered to 67 sophomore dental students in that course. Additionally, the impact of the source of the information, either lecture or text, on the accuracy of the response was studied. Data were collected from a total of 177 students during the three years of the study. Item analysis provided a difficulty index, and a discrimination index from the top 27 percent minus the lower 27 percent divided by 100. Kuder Richardson 20 was computed for each test, with values ranging from 0.59 to 0.68. Although it had been expected that short answer items would be the more difficult and best discriminators, the percentages of difficult or discriminating items did not vary greatly for item type or source, and neither factor produced a consistent trend. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED290767

A COMPARISON OF ITEM TYPE AND SOURCE ON DIFFICULTY AND DISCRIMINATION ABILITY

DIANA M. LANCASTER
ROBERT E. BARSLEY
CHARLES H. BOOZER

LSU SCHOOL OF DENTISTRY

PRESENTED AT THE ANNUAL MEETING OF MID-SOUTH EDUCATIONAL RESEARCH ASSOCIATION IN MOBILE, ALABAMA, NOVEMBER 13, 1987

TM 011 007

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Diana Lancaster

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

A Comparison of Item Type and Source on Difficulty and Discrimination Ability

Diana Lancaster, Robert Barsley, Charles Boozer
LSU School of Dentistry

The purpose of the present study was to compare difficulty and discrimination ability between multiple choice and short answer items. An additional consideration was to determine the effect of the source of the information for the questions - from lecture presentation, from text material or presented both in lecture and text.

Examinations for the internal medicine course at Louisiana State University School of Dentistry were constructed with about equal numbers of short answer and multiple choice items. The items were designated as having been presented in lecture, taken from the text, or presented in both lecture and text. Data was collected for three years. Item analysis provided a difficulty index (percent correct) and a discrimination index (top 27% minus lower 27% divided by 100). Data for the students over the three years were pooled for all like items. Kuder Richardson 20 was computed for each test and values ranged from .59 to .68.

It was expected that short answer items from the text would be the most difficult and best discriminators. However, percentages of difficult items or discriminating items did not vary greatly for either item type or source. Neither of these factors produced a consistent trend in performance.

INTRODUCTION

Test construction is of concern to dental educators because assessment of performance is an important aspect of the learning process.[1] The best method for assessing progress is debatable. The advantages and disadvantages of each type of item with respect to ease of construction, grading, and level of information tested for in both multiple-choice and short-answer questions have been explored.[2,3,4] The appropriateness of a particular item construction to the type of material (e.g., using multiple choice to test English material) has also been considered.[5] McClosky and Holland compared essay and multiple-choice questions, using scores for medical students on an examination in physiology. Their findings suggest better performance on multiple choice, but essay performance improves when cues are given.[6]

The purpose of this present study is to compare the difficulty of multiple-choice and short-answer items and their discrimination ability for student performance. An additional consideration is to determine whether the source of the information for the question - lecture presentation or text material - has an impact on the accuracy of responses. The following questions are considered:

1. Is there a difference between the multiple-choice and short-answer format in student test performance?
2. Does the source of the question - lecture or text have an effect on test performance?
3. If differences exist, are there implications for test construction and/or teaching strategies?

METHOD

The midterm and final exams for Internal Medicine at Louisiana State University School of Dentistry were constructed to allow comparison of the type of question (multiple choice or short answer) as well as the source (lecture or text). Items were written in each format and designated as having been presented in lecture material, taken from the text, or both.

The examinations were administered to 67 sophomore dental students in the Internal Medicine course. Responses were coded as correct or incorrect for both multiple choice and short answer items to allow for item analysis to be performed. The item analysis program provides the following:[3]

1. A difficulty index (percent who got the item correct).
2. The percentage of correct and incorrect responses for the top 27%, middle 46% and lower 27% of the scores.
3. A discrimination index (the top 27% minus the lower 27% divided by 100).
4. A point - biserial correlation coefficient.

The midterm consisted of 70 items - 34 multiple choice and 36 short answer. The final exam, which had 80 items, consisted of 40 short answer and 40 multiple choice.

There was some difficulty in constructing items from material which was presented in the text only consequently there are fewer of these items. The instructors believed these

items would be too specific and therefore, make the test unfair. The breakdown of items according to source is as follows: Midterm - 14 text, 14 - lecture, and 42 - lecture and text; final - 2 text, 42 - lecture, and 36 lecture and text. It was felt that the preponderance of items should be those covered in both lecture and text material.

The difficulty and discrimination indices were used to determine the overall best discriminators and most difficult items. Difficult items were defined as those that 50% or less answered correctly.[3] A good discriminator was defined by an index of .30 or greater.

After obtaining item analysis information, items were reviewed based on the criteria for difficulty and discrimination. Items which met either criterion were then grouped, based on type (multiple choice or short answer), or source (lecture or text). Frequencies were calculated for each type and each source to determine if any differences existed.

Since few items met the criteria, it appeared this might not be the best technique for analysis. Further, although there was some variability, no particular type of item appeared to be more difficult or a better discriminator. It was decided that the test would be given for the next 2 years and that data would be added to the analysis.

In order to provide test security, not all items were repeated on each test, some items were added and some were deleted. Further, topic outlines changed some what from year to year so that material that was on the midterm one year was on the final the next year.

In the three years the test was administered to a total of 177 students. There were a total of 92 items which remained the same over all three years. Due to the fact that some items were moved from the midterm to the final for the overall analysis all items were combined. Item analysis were also done for each of the tests separately. Means and standard deviations of scores were calculated and compared for the midterms and finals. The means were significantly different between the midterm and the final, with the final grades always being higher. The decision was made to combine the data as it was thought this difference wouldn't adversely affect the results.

Item analyses were performed for the 92 items for all the 177 students combined. The items were grouped according to type and source. They were also grouped based on combining the type and source eg. text/short answer and text/multiple choice. The values for difficulty and discrimination ability were analyzed separately. Means and standard deviations were calculated for each category. The criteria of good discriminator (,3) or difficulty (50% correct) were dropped as too few items met these. A t-test was used to compare difficulty between the multiple choice and short answer and discrimination ability

between these. Analysis of variance was used to compare differences among the sources and for the combined categories.

Kuder Richardson 20 was computed as a measure of reliability for each test separately and for the items combined.

RESULTS

The mean for the combined items was 78.8 and the standard deviation = 10.6. For the examinations individually the means ranged from 78 (sd = 6.7) on the 1984 midterm to 87.8 (sd = 5.2) on the 1986 final.

The values for KR 20 ranged from .59 to .68 on the individual tests and for the items combined KR 20 = .86. The reliability was improved by increasing the number of items, however, in general, the tests were moderately reliable.

The means and standard deviations for the items based on type and source of items are reported in Table 1. Difficulty and discrimination index data are reported separately. Analysis of variance was used to compare the means between the sources - lecture, test and lecture/text for difficulty and discrimination indices separately. Item types (multiple choice and short answer) were compared using t-tests.

For item source, difficulty F = .21, df = 2, 89 and for discrimination index F= .85, df = 2, 89. This indicates that there were not significant differences among item sources with

TABLE 1

MEANS AND STANDARD DEVIATIONS FOR ITEMS BY TYPE AND SOURCE

	Item Source			Item Type	
	Text	Lecture	Lecture/Text	Short Answer	Multiple Choice
Difficulty					
$\bar{X}$ =	77.63	80.39	78.09	77.68	79.88
sd =	18.05	12.42	17.37	15.7	15.9
n =	11	29	52	47	45
Disc. Index					
$\bar{X}$ =	.22	.26	.23	.24	.24
sd =	.11	.14	.12	.12	.13
n =	11	29	52	47	45

TABLE 2

MEANS AND STANDARD DEVIATIONS FOR ITEMS BY CATEGORIES COMBINED

	Text		Lecture		Lecture/Text	
	SA	MC	SA	MC	SA	MC
Difficulty						
$\bar{X}$ =	79.4	74.5	81.5	79.3	75.5	81.3
sd =	18.2	20	10.2	14.1	17	17.6
n =	7	4	13	16	28	24
Disc. Index						
$\bar{X}$ =	.20	.25	.26	.27	.25	.22
sd =	.12	.10	.15	.14	.11	.12
n =	7	4	13	16	28	24

regard to difficulty or discrimination. For item type - difficulty $t = .02$, $df = 90$ and for discrimination index $t = .69$, $df = 90$. This indicates that there was no significant difference between item type.

The items were combined with regard to type and source and the means and standard deviations are reported in Table 2. Analysis of variance was used to determine if there were significant differences among the six categories. For difficulty, $F = .49$, $df = 5, 86$ and for discrimination index, $F = .54$, $df = 5, 86$. This indicates that there were no significant differences among the groups considering type and source of the items together.

DISCUSSION

Based on item analysis information, a comparison of item type and source revealed no significant differences with respect to difficulty and discrimination. It was anticipated that short answer items from text material would be the most difficult and the most effective discriminators with respect to student performance. The results did not support this assumption. Short answer and multiple choice questions were about equal with regard to difficulty and discrimination ability. The mean for the lecture based items was higher but this difference was not significant. The source of the material did not seem to have any significant impact on test performance.

For the items generated on the material in the Internal Medicine course, both types of items performed equally as well.

Having material presented in both lecture and text did not improve performance on those items.

CONCLUSIONS

Although the information obtained in this study is situation specific, some statements may be made for consideration.

Recall of information which is required in the short answer format is generally considered to be more difficult than recognition as in the multiple choice format; however, they were of about equal difficulty for this material. The difficulty of the material being tested may be more important than the item format.

When considering item source, material covered in both lecture and text would seem to increase the likelihood it will be learned. However, again, these items were of about equal difficulty with those from lecture or text alone. The small number of items that were from the text alone limits the interpretation. Whether or not presenting material in more than one format enhances learning could be investigated further.

The only significant finding appears to be the fact that performance improved on the final which may indicate that the students improved in their test taking ability and deciding what should be learned.

References

1. Slavicek, P. A Psychologist's View of Knowledge Testing and Evaluation in the Process of Education and Instruction. Acta Universitatis Carolinae Medica. 29(3/4) p 127-129, 1983.

2. Bloom, B.S. et. al. ed. Handbook on Formative and Summative Evaluation of Student Learning. New York, McGraw-Hill Book Co. 1971 pp 145-149.

3. Thorndike, R.L. ed. Educational Measurement, 2nd ed. Washington, D.C. American Council on Education, 1971 pp 62-63, 131-145.

4. Hodge, G.T. How to Make Tests More Effective. Educational Directions 5(2):9-12, 1980.

5. Bamberg, B. Multiple Choice and Holistic Essay Scores: What Are They Measuring? College Composition and Communication 33(4):404-406, 1982.

6. McClosky, D. and R. Holland. A Comparison of Student Performances in Answering Essay-Type and Multiple-Choice Questions. Medical Education 10(5):382-385, 1976.